

BILS Proteomics

Pipeline for Swedish Mass Spectrometry Data

Background

- Mass Spectrometry (MS) based proteomics generates large quantities of data in different formats
- Diversity of analysis pipelines
- Lab handling of data is problematic
- Little data is published in reusable format

BILS

Bioinformatics
Infrastructure
for Life Sciences



BILS – MS proteomics

- Build infrastructure for reliable long-term storage of raw MS data.
- Give Swedish researchers access to bioinformatics tools for proteomics.
- Help with publication of proteomics experiments.
- Collaborate with SNIC for storage and computing resources

BILS

Bioinformatics
Infrastructure
for Life Sciences



Demands for BILS solution

- Raw data files should be automatically uploaded to Swegrid storage.
- Files needs to be annotated, otherwise useless for others and difficult to track.
- Files not accessible by other people then the submitting lab before release.
- Data should be in a format readable by other researchers without access to instrument software (mzML)

BILS

Bioinformatics
Infrastructure
for Life Sciences



Implemented solution

- Automated scripts for conversion and annotation of new MS raw data
- Upload of files to Swestore with access only for submitter. Registration of files and annotations in local web application (Proteios).
 - Files are in Proteios found under their respective user and project.

BILS

Bioinformatics
Infrastructure
for Life Sciences





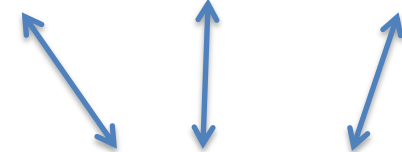
Web browser access and analysis
of own data only



Proteios

Software Environment

*Welcome to proteios at
www.proteios.org.*



SWESTORE NATIONAL STORAGE



Public access to
released raw
data



Step 1 – File conversion

- Script looking for list of files to convert in spreadsheets.

Name	Project	User	Sample name	Sample organism	Inj vol	Diution factor
FL_100524_red_1	Label test	fredrik	red	Mus musculus	2	5
FL_100524_blu_2	Label test	fredrik	blue	Mus musculus	2	5

- Conversion to mzML (and MGF) using Proteowizard. Generation of an info file for each file.

Step 2. Upload to Swestore

- Script searching upload folder for .info files.
- Upload raw data and mzML files to Swestore

File organisation:

Lab name -- Instrument – type -- month – file

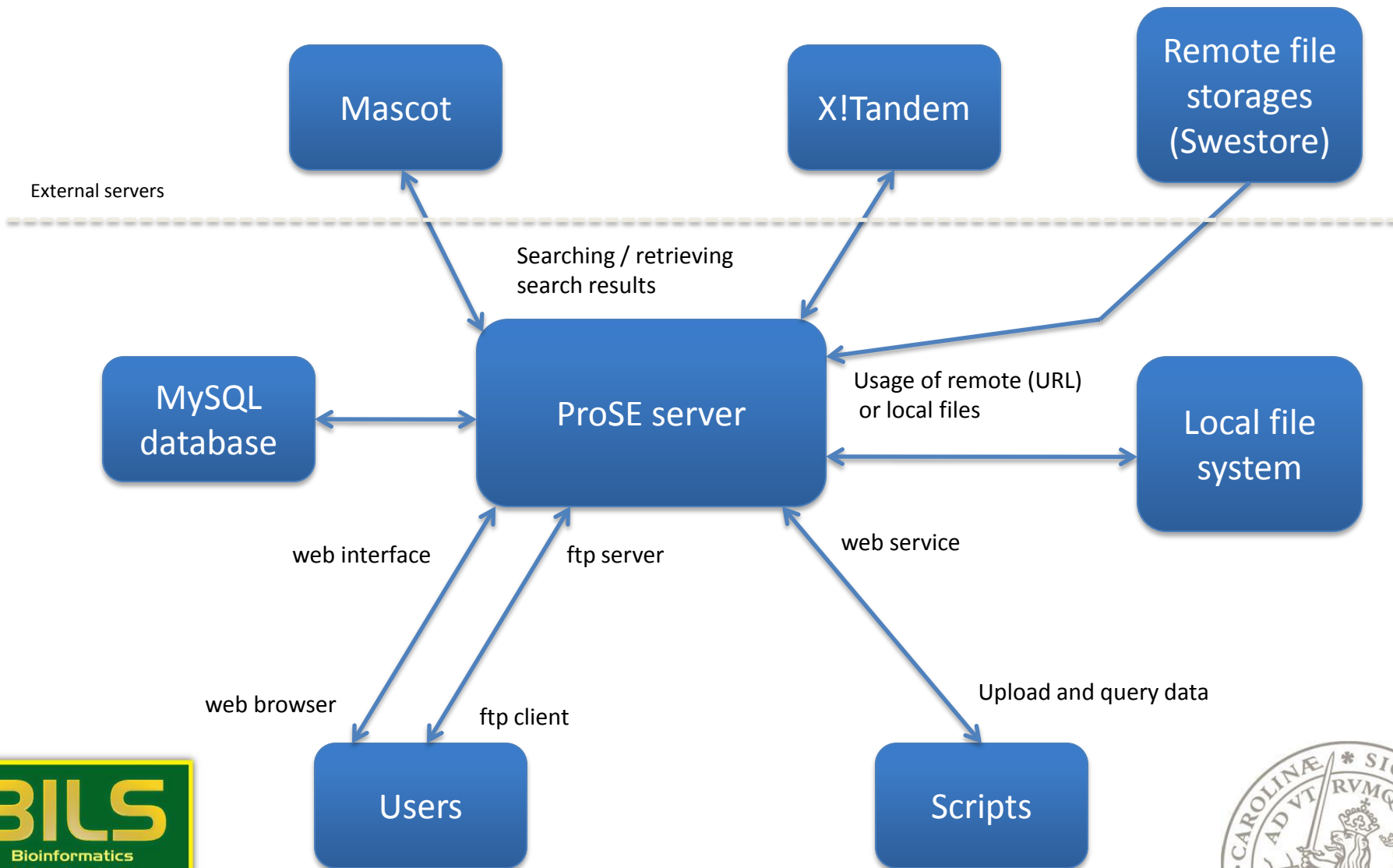
-> Registration of files in Proteios. Upload of MGF files to Proteios.

BILS

Bioinformatics
Infrastructure
for Life Sciences



Proteios SE – analysis system



Proteios interface - Benefits

- User- and project-based access to files on Swestore.
- Batch usage of analysis tools (Mascot, X!Tandem, OpenMS, msInspect, etc.)
- Quantitative analysis (isobaric labels, label free)
- Automatic linking of samples and search results
 - Quantitative reports with fixed FDR
- Export of projects to repository (PRIDE / ProteomeXchange).

BILS

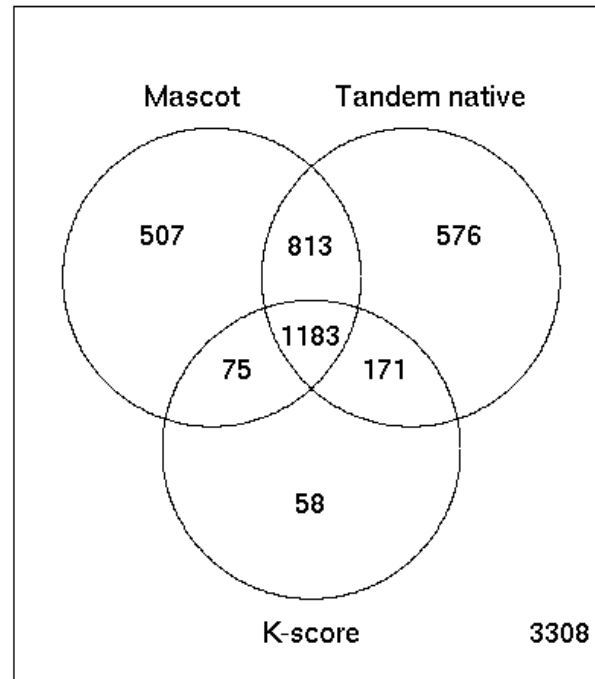
Bioinformatics
Infrastructure
for Life Sciences



Example: Combining results from different search engines

- Complementary results from different search engines (Kapp et al, Proteomics 2005)

In-house example:



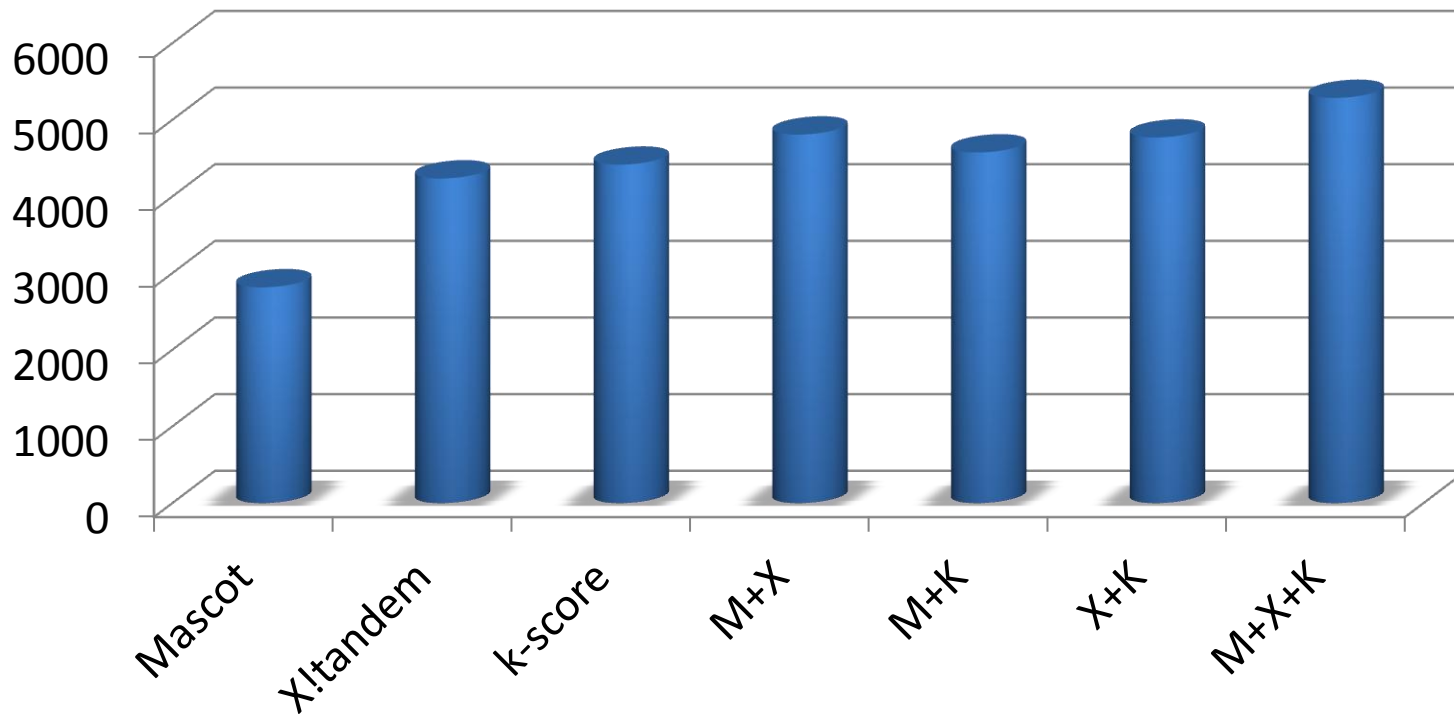
BILS

Bioinformatics
Infrastructure
for Life Sciences

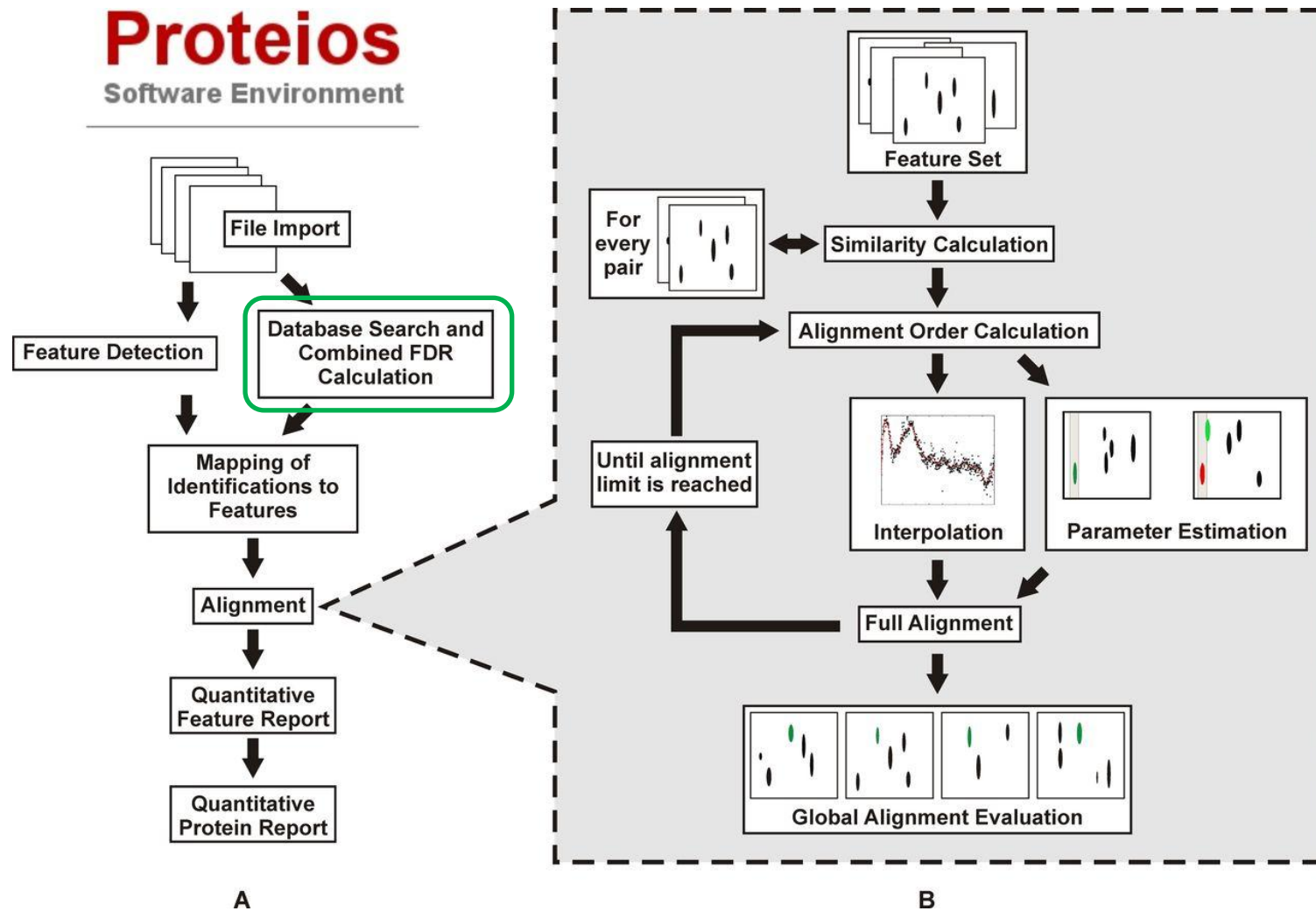
Peptide identifications

Orbitrap using Proteios SE

CID Peptide IDs at FDR 0.01



Integrated quantification workflow



Sandin M et al. Mol Cell Proteomics 2013;12:1407-1420



Publishing help

- Deposition of raw data – publish by changing access rights to files.
- Keep track of parameters and settings in accordance with MIAPE (Minimum Information About a Proteomics Experiment)
- Export to repository in PRIDE repository XML format
- ProteomeXchange submission files generation

BILS

Bioinformatics
Infrastructure
for Life Sciences



Published on Swestore

[Mol Cell Proteomics](#). 2012 Aug;11(8):342-54. Epub 2012 Apr 27.

Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry.

[Olsson N](#), [James P](#), [Borrebaeck CA](#), [Wingren C](#).

[J Proteome Res](#). 2012 Jul 6;11(7):3766-73. Epub 2012 Jun 14.

Automated selected reaction monitoring software for accurate label-free protein quantification.

[Teleman J](#), [Karlsson](#)

[Mol Cell Proteomics](#). 2011 Oct;10(10):M110.003962. Epub 2011 Jun 14.

Proteomic analysis and discovery using affinity proteomics and mass spectrometry.

[Nat Commun](#). 2012 Jul 31;3:976. doi: 10.1038/ncomms1975.

Proteomic screen reveals Fbw7 as a modulator of the NF- κ B pathway.

[Arabi A](#), [Ullah K](#), [Branca RM](#), [Johansson J](#), [Bandarra D](#), [Haneklaus M](#), [Fu J](#), [Ariès I](#), [Nilsson P](#), [Den Boer ML](#), [Pokrovskaja K](#), [Grandér D](#), [Xiao G](#), [Rocha S](#), [Lehtiö J](#), [Sangfelt O](#).

[Mol Cell Proteomics](#). 2012 Jul;11(7):M112.016998. Epub 2012 Apr 12.

Tumor proteomics by multivariate analysis on individual pathway data for characterization of vulvar cancer phenotypes.

[Sandberg A](#), [Lindell G](#), [Källström BN](#), [Branca RM](#), [Danielsson KG](#), [Dahlberg M](#), [Larson B](#), [Forshed J](#), [Lehtiö J](#).

BILS

Bioinformatics
Infrastructure
for Life Sciences



Current Status and outlook

- Labs from Stockholm, Lund, Uppsala and Gothenburg using the pipeline.
- Developments: Tight integration with ProteomeXchange. Further tool integration. Label free quantification workflow.
- Searchable database of published Swedish experiments

BILS

Bioinformatics
Infrastructure
for Life Sciences

